

REMARKS

In the April 28, 2005 Office Action, the Examiner noted that claims 1-20, 22-24, 26, 27, 29-31 and 33-43 were pending in the application; rejected claims 1, 2 and 10-17 under 35 USC § 102(b); and rejected all of the pending claims under 35 USC § 102(a). In rejecting the claims, the Examiner cited an article entitled "Content-Based Retrieval for Music Collections" by Yuen-Hsien Tseng (hereafter Tseng) and an article entitled "Automatic Audio Content Analysis" which was identified as being authored by "S. Pfeiffer," however the article with this title that was attached to the April 28, 2005 Office Action was authored by three people, the first of whom was identified as Silvia Pfeiffer and therefore will be referred to below as Pfeiffer et al. Claims 1-20, 22-24, 26, 27, 29-31 and 33-43 remain in the case. The Examiner's rejections are traversed below.

Rejections under 35 USC § 102(b)

In item 3 on page 3 of the Office Action, claims 1, 2 and 10-17 were rejected under 35 USC § 102(b) as anticipated by Pfeiffer et al. In making this rejection, it was asserted that Pfeiffer et al. "teaches simulating (i.e. modeling human perception of music)" (Office Action, page 3, lines 1-2) in sections 2.2 and 3.2-3.3.2 on pages 22-24. In section 2.2 which is part of the section entitled "BASIC PROPERTIES OF AUDIO", Pfeiffer et al. describes "two methods of simulating human auditory perception" using a computer which either "model the human auditory system and every detail that is known, or ... make black box models of the processes occurring in the human auditory system" (page 22, left column, lines 35-39). The first technique is disregarded by Pfeiffer et al. and the remainder of section 2.2 discusses the use of loudness and "sound history."

In Pfeiffer et al., "sound history" is defined as the "profile of the loudness the human has perceived in the past (for example during the last 2 minutes)" (page 22, right column, lines 31-33) combined with an "'intersubjective' loudness measure" (page 22, right column, line 36) which is "derived from the environment" (page 22, right column, line 34). In addition to such loudness "sound history" which Pfeiffer et al. also refers to as "Volume analysis" (page 22, right column, line 48), Pfeiffer et al. describes a toolbox of "common operators of digital audio analysis ... [in which t]he indicators the toolbox contains ... are ... Frequency analysis, Pitch analysis, Onset and Offset, Frequency transition maps, Audio segmentation, Fundamental frequency analysis and Beat analysis." (page 22, right column, line 41 to page 23, left column, line 4). In addition, Pfeiffer et al. describes a method in which "frequencies are filtered first in a perception-simulating analysis. The filter ... computes the response a specific nerve cell of the auditory nerve will

produce" (page 23, left column, lines 13-15). It is submitted that simulating a nerve cell does not teach or suggest "a computational model of human perception of a descriptor of music" (claim 1, lines 1-2) or make it possible to provide a "computed descriptor for each recording [that] most closely matches perceptions reported for the recording by one or more human listeners" (claim 1, last 2 lines).

The preamble and operation (b) of claim 10 include a limitation similar to the one quoted in the preceding paragraph from claim 1 and claims 3, 5, 6, 10 and 29 include a limitation similar to the limitation at the end of claim 1 quoted above. Therefore, it is submitted that claims 1, 3, 5, 6, 10 and 29 and claims 2, 4, 7-9 and 11-17 which depend therefrom, patentably distinguish over Pfeiffer et al. due to the failure of Pfeiffer et al. to disclose a model or descriptor of human perception as recited in these claims.

Next, the Office Action asserted that Pfeiffer et al. discloses "extracting numeric parameters from an electronic representation of musical recordings" (Office Action, page 3, lines 3-4) at sections 3.3 and 4.1 with reference to Figs. 5 and 6. These portions of Pfeiffer et al. describe "first ... content-based audio segmentation ... to distinguish between music, speech, silence and other sound sequences" (page 23, right column, lines 17-19). Clearly, this does not involve "extracting ... at least two numeric parameters" (e.g., claim 1, lines 3-4). Next, "[t]he second step is the classification of the segments into speech, music, silence and other sounds" as stated in the sentence spanning pages 23 and 24. No description of how this is accomplished has been found in Pfeiffer et al., only an "idea ... to distinguish between music and other sounds by analyzing the spectrum for 'orderliness': tones and their characteristic overtone pattern do not appear in environmental sounds, neither is a rhythmic pattern present" (page 24, left column, lines 12-16). It is unclear whether the Examiner is asserting that "tones and their characteristic overtone pattern" or "rhythmic pattern" are being interpreted as either "numeric parameters" (claim 1, line 4) or "a single number representing the description for that recording" (claim 1, line 6). However, it is submitted that nothing in Pfeiffer et al. suggests "combining the numeric parameters with a weighting for each parameter to compute a single number representing the descriptor for that recording" (e.g., claim 1, lines 5-6) for the reasons discussed below.

Section 3.3.2 of Pfeiffer et al. discusses "a fundamental frequency (fuf) determination of the chords as a first step toward note analysis" (page 24, right column, lines 29-31). According to Pfeiffer et al., "[t]he sequence of fuf's in a piece of music ... is one of the parameters most important in determining the structure of a piece of music" (page 24, right column, lines 32-35). Thus, it appears that Pfeiffer et al. teaches extraction of a sequence of fufs as a parameter.

However, this is the only parameter for which any description has been found of being used by the method described in Pfeiffer et al. In fact, Fig. 6 which was cited by the Examiner twice uses the term "fuf recognition" implying that the recognition of music is based solely on the fundamental frequency or fuf.

Furthermore, section 4.1 describes "[t]he compression of a piece of music into a sequence of fundamental frequencies (fufs) ... [as] a means to produce a characteristic signature of music pieces ... for audio retrieval, where music must be recognized" (page 25, right column, lines 6-9). At the end of section 4.1, the authors acknowledged that they "cannot yet produce a characteristic signature based solely on fuf indicators" (sentence spanning pages 26 and 27). Therefore, Pfeiffer et al. suggests combining "fuf indicators ... with FFT indicators ... [as being] more reliable at the moment" (page 27, left column, lines 1-2). However, no description of extracting Fast Fourier Transform (FFT) indicators has been found in Pfeiffer et al. Therefore, Pfeiffer et al. is a non-enabling reference with respect to extracting more than one parameter from music and at most only suggests doing so.

The Office Action asserted that the fuf recognition illustrated in Figs. 6-8 constitutes teaching "a model that considers the likeness (i.e., similarities) between the extracted representation of musical recordings" (Office Action, page 3, lines 7-8). This statement suggests a failure to understand the distinction between music that is "similar" and music that is found to be the "same." As indicated by the bottom two block in Fig. 6, the purpose of the comparison based on signatures formed by fuf sequences in the method described by Pfeiffer et al. is to determine whether music in a commercial extracted from a broadcast (the right side of Fig. 6) is "found" to be the same as one of the signatures in the Commercials Database on the left side of Fig. 6, or whether a "new piece" is in the commercial being broadcast.

The only description that has been found of what was used to analyze music in the tests performed by the authors of Pfeiffer et al. appears in the only full paragraph in the left column of page 26. This description is at such a high level it would not enable a person of ordinary skill in the art to duplicate the test system. Therefore, it is submitted that Pfeiffer et al. is a non-enabling reference.

Furthermore, as described in the only full paragraph in the left column of page 26, the music recognition process described in Pfeiffer et al. uses "three classes of indicators ... evaluated **separately**" (page 26, left column, lines 17-18, emphasis added). For each entry in the database there are "10 FFT analysis indicators without use of windows, 10 FFT analysis indicators using a Hanning window and 10 fuf analysis indicators" (page 26, left column, lines 15-17).

If the Examiner has assumed that the FFT indicators should be considered parameters, this would mean that the database contains ten different values for three different parameters. However, the description of how these classes of indicators are used states that each is used **separately**; thus, there is no teaching or suggestion in this paragraph of "combining the numeric parameters with a weighting for each parameter to compute a single number representing the descriptor for that recording" (e.g., claim 1, lines 5-6). Rather,

[t]hese 30 indicators are also calculated for a queried piece of music and then compared to each respective indicator of the pieces in the database, resulting in a similarity percentage. The highest similarity percentage determines the entry that is identified by a single indicator. Then, the indicator results are accumulated within the four (sic) classes.

(page 26, lines 19-24). Thus, each accumulated indicator results is a single numeric value, that does not constitute "a descriptor for that recording" (e.g., claim 1, line 6), because it is not for a single recording, but rather represents the comparison between a recording in the database with the queried piece and is used to determine whether a match has been found. For the above reasons, it is submitted that claim 1 and claim 2 which depends therefrom are not anticipated by Pfeiffer et al.

In the Office Action, it was asserted that Pfeiffer et al. "discloses extracting numeric parameters from recordings and the use of weighting parameters" (page 3, lines 8-9) at section 4.2 with reference to Tables 2 and 3. First, it is noted that section 4.2 is entitled "Violence Detection" and describes that the authors "extracted shots, explosions and cries out of audio tracks manually" (page 27, right column, lines 10-11). Thus, section 4.2 of Pfeiffer et al. describes a different use of frequency analysis of audio than section 4.1. There is not even a suggestion in Pfeiffer et al. that section 4.2 relates to music, let alone, "human perception of a descriptor of music" (claim 1, lines 1-2) or "electronic representations of musical recordings" (claim 1, line 3). Since some form of the word "music" is used in all of the independent claims, it is submitted that section 4.2 of Pfeiffer et al. is not relevant to the present invention.

Furthermore, what is described in section 4.2 of Pfeiffer et al. is the use of the indicators: "loudness, frequency, pitch, onset, offset, [and] frequency transitions" (page 27, right column, lines 24-25) for each of which "minimum, maximum, mean, variance and medium statistics" (page 27, right column, lines 25-26) are calculated and are then linearly combined using heuristically determined weights shown in Table 2. The heuristic process used to determine the weights is not described in Pfeiffer et al. other than that "in most cases the correlation between mean and variance is higher than that between mean and maximum" (page 27, right column,

lines 30-31) and that the "weights differ from event to event" (page 28, left column, lines 2-3) as illustrated in Fig. 3, where "event" refers to one of a shot, cry or explosion. No teaching or even suggestion has been found in Pfeiffer et al. of applying the weighting of either indicators or statistical elements used in an attempt to identify shots, cries and explosions to the identification of music as discussed in section 4.1. For the above additional reasons, it is submitted that claim 1 and claim 2 which depends therefrom patentably distinguish over Pfeiffer et al.

While claim 1 is directed to defining a descriptor by combining at least two weighted numeric parameters, claim 10 is directed to "a computational model of human perception of **likeness** between musical recordings" (claim 10, lines 1-2, emphasis added). As should be clear from the specification, "likeness" is distinct from "sameness" or determining whether two recordings obtained in different ways are the same recording or the same piece of music. Therefore, it is submitted that Pfeiffer et al. is not relevant to the method recited in claim 10, since it is directed to identifying whether musical recordings are the same.

Furthermore, claim 10 recites "receiving from one or more human listeners who compares pairs of musical recordings an indication of the human's perception of likeness for each compared pair of recordings" (claim 10, lines 5-7). As discussed above, Pfeiffer et al. uses humans to identify whether sounds represent shots, cries or explosions. There is no suggestion that they are similar sounding cries or similar sounding explosions, or that this technique could be applied to musical recording. As also discussed above, what heuristic process was used to develop the weightings of statistical elements and indicators used to identify shots, cries and explosions, was not disclosed in Pfeiffer et al. Moreover, there is no suggestion that the way the weights were heuristically determined involved "a weighting for each parameter comparison number to compute a single difference number representing the difference between ... two recordings" (claim 10, lines 13-14), where the "parameter comparison" results from "comparing each numeric parameter of one recording in the pair with the corresponding parameter of the second recording of the pair using an algorithm which produces a parameter comparison number representing the parameter comparison" (claim 10, lines 8-11). For all of the above reasons, it is submitted that claim 10 and claims 11-17 which depend therefrom patentably distinguish over Pfeiffer et al.

Furthermore, claim 13 recites details of how "a single number representing a descriptor for ... [a] recording" (claim 13, line 4) is formed and used in comparing the numeric parameters. It is submitted that the details recited in claim 13 are not taught or suggested by Pfeiffer et al. Therefore, claim 13 and claim 17 which depends therefrom further distinguish over Pfeiffer et al.

Rejections under 35 USC § 102(a)

In item 4 on pages 4-6 of the Office Action, all of the pending claims were rejected under 35 USC § 102(a) as anticipated by Tseng. In rejecting claims 1 and 10-13, it was asserted that Tseng describes "a model formed from the perception of the music ... as perceived by human subjects inclusive of extracting numeric parameters (i.e., descriptors) from an electronic representation of musical recordings" (Office Action, page 4, lines 3-6) in the Abstract and Sections 1-4 with reference to Fig. 1 and Tables 1-4. However, the only thing that is extracted from music and is used by the method described in Tseng is information related to notes to create "a pitch profile" (e.g., Abstract, line 3) for a recording and "n-note indexing" (Abstract, line 4). An example of how a pitch profile is created is described in the last full paragraph on page 178.

As in the case of Pfeiffer et al., the purpose of the method described in Tseng is to identify a specific song, not determine whether two different recordings are similar, as in the case of the present invention. Unlike Pfeiffer et al., where a musical recording is available for querying purposes so that the same extraction technique can be used on the unidentified recording, Tseng is directed to a system in which users input note information which is then matched with the pitch profile using an n-note index to locate a musical recording having that pitch profile. The system described in Tseng uses a "key melody extraction module that extracts representative and memorable melodies from the music collection for query suggestion and effective retrieval" (page 177, left column, lines 14-16). As illustrated in Fig. 1, the method described in Tseng starts with a Musical Instruments Digital Interface (MIDI) collection of music from which melodies are extracted in the form of melody strings. From these melody strings, "key melody strings" are extracted although how this is done and exactly what constitutes a "key melody string" is not described. The key melody strings are used to reconstruct MIDI files corresponding thereto using the MIDI collections.

In Section 4 of Tseng, several experiments using differing types of pitch encoding and indexing are described. Six retrieval modes were tested which are listed at the top of the right column on page 180 with five of the modes numbered and the first mode unnumbered on lines 1-2. Twenty queries were generated by eight individuals which are listed in Table 2 on page 181. As described below Table 2 in the left column on page 181, two queries were generated for each melody query, one in absolute pitch using standard notation and the other in relative pitch using simplified notation. To enable users to formulate queries, "they are allowed to listen to the music and make their queries based on their perception of the tunes" (page 180, right column, lines 20-21).

It is clear from Table 3 on page 181 of Tseng that some of the queries resulted in multiple matches, with the number of matches varying depending upon mode. For example, mode 4 produced at least three hits and as many as 421 out of 444 records in the database. This does not mean that any human would consider almost all of the songs in the database as perceptually similar, but rather that either the melody strings or the key melody strings extracted from most of the songs in the database include a sequence of notes matching the key-invariant form of the query listed in Table 2 as "song" 11.

As noted above, Tseng states that "melodies, rhythms and chords information can be extracted" (page 178, right column, lines 17-18) from MIDI files; however, only one of these, i.e., notes that form melodies, are extracted in the method described in Tseng. Thus, the method described in Tseng does not extract "at least two numeric parameters" (e.g., claim 1, lines 3-4). This limitation is recited in all of the independent claims except claim 26. Therefore, it is submitted that all of the claims, except claim 26, patentably distinguish over Tseng for this reason.

In the Office Action, it was asserted that Tseng describes "the use of weighting parameters ... to compute (calculate) the correlations between recorded sections (i.e., the stored numerical descriptors) and subsequently adjusting the weighting based on human perception" (Office Action, page 4, lines 10-12) in paragraph 2 of the left column on page 179 and on pages 180-181 with reference to tables 1-4. No indication of weighting has been found in tables 1-4 or anywhere on pages 180-181. The Examiner is respectfully requested to identify with more specificity where weighting has been found in section 4 of Tseng if this rejection is maintained.

The term "weight" is used in the first full paragraph on page 179 in the statement that "[t]he weight of each m-note in queries is given by $2(m-1)+1$ while the weight of each of m-note in documents takes the value of 1 or 0, indicating its presence in the documents or note" (page 179, left column, lines 20-23). No explanation of how this weight is used has been found in Tseng. Furthermore, it is submitted that this statement does not suggest "combining the numeric parameters with a weighting for each parameter to compute a single number representing the descriptor for that recording" (e.g., claim 1, lines 5-6). In fact, no suggestion has been found that the system described in Tseng generates any sort of descriptor for any recording, let alone creating such a descriptor by combining weighted parameters. As discussed above, Tseng is directed to retrieving MIDI files by supplying a sequence of notes using an index of two or three notes using melodies or "key" melodies extracted from MIDI files. Due to the lack of teaching regarding the use of weights in Tseng, there is no suggestion of "adjusting the weighting for the parameters" by any operations, certainly not as recited on the last three lines of claim 1.

For the above reasons, it is submitted that claim 1, as well as claims 2, 34 and 35 which depend therefrom, patentably distinguish over Tseng.

Using language similar to that in claim 1, claim 10 recites "extracting ... at least two numeric parameters" (claim 10, lines 3-4), "combining the parameter comparison numbers with a weighting for each parameter comparison number to compute a single difference number representing the difference between the two recordings of the pair" (claim 10, lines 12-14) and "adjusting the weightings" as recited on the last 3 lines of claim 10. Therefore, it is submitted that claim 10, as well as claims 11-17, 42 and 43 which depend therefrom, patentably distinguish over Tseng for at least the reasons discussed above with respect to claim 1.

Claims 3, 5-7 were addressed in the only full paragraph on page 5 of the Office Action. Using language similar to that in claim 1, claims 3, 5 and 6 recite "extracting ... at least two numeric parameters" (claim 3, lines 4-5 and 9-10; claim 5, lines 5-6 and 10-11; and claim 6, lines 8-9 and 13-14), "combining the numeric parameters with a weighting for each parameter to compute a single number representing the descriptor for that recording" (claim 3, lines 6-7; claim 5, lines 7-8; and claim 6, lines 10-11) and "adjusting the weightings" as recited on the last 3 lines of claims 3 and 5 and lines 17-19 of claim 6. Therefore, it is submitted that claims 3, 5 and 6, as well as claims 4, 7-9 and 36-42 which depend therefrom, patentably distinguish over Tseng for at least the reasons discussed above with respect to claim 1.

Furthermore, claim 3 recites "generating a data record associated with a music recording, the record comprising two or more scalar descriptors, each descriptor numerically describing the recording of music" (claim 3, lines 1-3) and claim 5 recites "a computer extracted data record associated with a music recording, the record comprising: two or more scalar descriptors, each descriptor numerically describing the recording of music" (claim 5, lines 1-4). Similarly, claim 6 recites "data records each comprising two or more scalar descriptors" (claim 6, line 5). It is submitted that even when melodies or key melodies are presented by a key-invariant string using only numerals, as taught by Tseng, such strings or the index entries derived from these strings are not "scalar" (e.g., claim 3, line 2) nor do they "numerically" (e.g., claim 3, line 2) describe a recording, because they are strings, not "a single number representing the descriptor for that recording" (e.g., claim 3, line 7). Furthermore, these strings are not created by "combining the numeric parameters" (e.g., claim 3, lines 6). For the above additional reasons, claims 3, 5 and 6 and claims 4 and 7-9 which depend from claim 3 and 6 further patentably distinguish over Tseng.

Claims 18-20 were also addressed in the only full paragraph on page 5 of the Office Action. Using language similar to that in claim 1, claim 18 recites "extracting ... at least two numeric parameters" (claim 18, lines 4-5). Although the term "descriptor" is not used in claim 18, it does recite "computing from the extracted parameters for each of a plurality of pairs of the recordings a number which represents the difference between the recordings of the pair" (claim 18, lines 6-7). In addition, claim 18 recites "assembling the computed difference numbers into a database where each computed difference is associated with the identifier for each of the two recordings from which the difference was computed" (claim 18, last 3 lines). The Office Action did not cite anything and nothing has been found in Tseng teaching or suggesting a database that contains such difference numbers. For at least the reasons discussed above with respect to similar limitations, it is submitted that claim 18, as well as claims 19, 20 and 22-24 which depend therefrom, patentably distinguish over Tseng.

Claim 26 was also listed at the beginning of the only full paragraph on page 5 of the Office Action. However, nothing in this paragraph refers to any teaching in Tseng of the limitation recited in the last three lines of claim 26, "searching a database containing computed difference numbers between the target recording and a plurality of other recordings for those recordings which have a small computed difference number from the target music recording." As discussed above with respect to claim 18, nothing has been found in Tseng suggesting the use of a database containing computed difference numbers as recited in claim 26. Therefore, it is submitted that claim 26 and claims, as well as claims 27, 29-31 and 33 which depend therefrom, patentably distinguish over Tseng.

Summary

It is submitted that the references cited by the Examiner, taken individually or in combination, do not teach or suggest the features of the present claimed invention. Therefore, it is submitted that claims 1-20, 22-24, 26, 27, 29-31 and 33-43 are in a condition suitable for allowance. Reconsideration of the claims and an early Notice of Allowance are earnestly solicited.

Finally, if there are any formal matters remaining after this response, the Examiner is requested to telephone the undersigned to attend to these matters.

Serial No. 09/556,086

If there are any additional fees associated with filing of this Amendment, please charge the same to our Deposit Account No. 19-3935.

Respectfully submitted,

STAAS & HALSEY LLP

Date: 10/28/05

By: Richard A. Gollhofer
Richard A. Gollhofer
Registration No. 31,106

1201 New York Avenue, NW, Suite 700
Washington, D.C. 20005
Telephone: (202) 434-1500
Facsimile: (202) 434-1501